applied sciences

MDPI

Article

Research on the Anchor-Rod Recognition and Positioning Method of a Coal-Mine Roadway Based on Image Enhancement and Multiattention Mechanism Fusion-Improved YOLOv7 Model

Xusheng Xue [1,2], Jianing Yue [1,2,*], Xingyun Yang [1,2], Qinghua Mao [1,2], Yihan Qin [1,2], Enqiao Zhang [1,2] and Chuanwei Wang [1,2]

[1] School of Mechanical Engineering, Xi'an University of Science and Technology, Xi'an 710054, China; xuexsh@xust.edu.cn (X.X.); 15700971803@163.com (X.Y.); maoqh@xust.edu.cn (Q.M.); qyh1484083485@163.com (Y.Q.); mikonzhang@gmail.com (E.Z.); wangchuanwei228@xust.edu.cn (C.W.)
[2] Shaanxi Key Laboratory of Mine Electromechanical Equipment Intelligent Detection and Control, Xi'an 710054, China
* Correspondence: 18235402330@163.com

Abstract: A drill-anchor robot is an essential means of efficient drilling and anchoring in coal-mine roadways. It is significant to calculate the position of the drill-anchor robot based on the positioning information of the supported anchor rod to improve tunneling efficiency. Therefore, identifying and positioning the supported anchor rod has become a critical problem that needs to be solved urgently. Aiming at the problem that the target in the image is blurred and cannot be accurately identified due to the low and uneven illumination environment, we proposed an improved YOLOv7 (the seventh version of the You Only Look Once) model based on the fusion of image enhancement and multiattention mechanism, and the self-made dataset is used for testing and training. Aiming at the problem that the traditional positioning method cannot guarantee accuracy and efficiency simultaneously, an anchor-rod positioning method using depth image and RGB image alignment combined with least squares linear fitting is proposed, and the positioning accuracy is improved by processing the depth map. The results show that the improved model improves the mAP by 5.7% compared with YOLOv7 and can accurately identify the target. Through the positioning method proposed in this paper, the error between the positioning coordinate and the measurement coordinate of the target point on each axis does not exceed 11 mm, which has high positioning accuracy and improves the positioning accuracy and robustness of the anchor rod in the coal-mine roadway.

Keywords: anchor-rod positioning; image enhancement; multiattention mechanism; visual target detection

check for updates

Citation: Xue, X.; Yue, J.; Yang, X.; Mao, Q.; Qin, Y.; Zhang, E.; Wang, C. Research on the Anchor-Rod Recognition and Positioning Method of a Coal-Mine Roadway Based on Image Enhancement and Multiattention Mechanism Fusion-Improved YOLOv7 Model. *Appl. Sci.* **2024**, *14*, 1703. https://doi.org/10.3390/app14051703

Academic Editor: Andrea Prati

Received: 31 January 2024
Revised: 18 February 2024
Accepted: 18 February 2024
Published: 20 February 2024

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Coal-mine production is mainly based on underground mining, and roadway development depends on tunneling operations. Therefore, anchor support is vital in coal-mine production [1,2]. Improving the efficiency of roadway tunneling has become one of the essential contents of coal-mine intelligence. At present, the main factor affecting the efficiency of roadway excavation is the intelligent level of drilling and anchoring operations. On 2 January 2019, the National Coal Supervision Bureau issued the 'Key R&D Directory of Coal Mine Robots'. On 25 February 2020, the National Development and Reform Commission and eight other ministries and commissions jointly issued the 'Guiding Opinions on Accelerating the Intelligent Development of Coal Mines,' which clarified the development goals and main tasks of coal mine drill-anchor robots. Drill-anchor robots are of great significance in solving the problem of supporting efficiency, but the positioning of drill-anchor robots is mainly done by manual operation, which cannot ensure the accuracy and

efficiency of fuselage positioning [3,4]. The positioning of drill-anchor robots can be solved by taking the supported anchor rods as a reference [5]. However, there are few studies on the identification and location of anchor rods, and the complex roadway environment increases the difficulty of anchor-rod identification [6]. Therefore, an in-depth study of the identification and positioning technology of the anchor rods in the roadway, especially in the environment of low and uneven illumination in coal-mine roadways, is of great significance for the precise positioning of drill-anchor robots [7].

Currently, most studies for object recognition in uneven lighting and low illumination environments are implemented based on machine vision. Aiming at the problem of image enhancement under low illumination, Cheng et al. [8] integrated the Swin v2 module and convolution module and adopted the attention mechanism of a multiscale module. A low illumination image-enhancement algorithm based on the Transformer model was proposed to realize the enhancement of low illumination images in a coal mine. Reference [9] used the MobileNetV2 feature-extraction network as the backbone network of DeepLabV3+ and adopted the channel attention (SE) and channel spatial attention (CA) mechanisms to improve the accuracy of coal-rock recognition. Zhang et al. [10] added a coordinate attention mechanism module to the backbone network of the YOLOv5s model and increased the clarity of the anchor-hole image through super-resolution reconstruction technology, which effectively improved the accuracy of recognition and classification. Hao et al. [11] added a convolutional block attention model into the YOLOv5 network to enhance the saliency of the target in the image, aiming at the problem that the target cannot be accurately identified due to the uneven illumination environment. Xu et al. [12] proposed an Attention Res-UNet method. U-Net is used for semantic segmentation and uses an attention mechanism to reduce the influence of uneven lighting conditions, which proves its recognition superiority in complex environments. Based on traditional YOLOv5s, Yang et al. [13] proposed an ODEL-YOLOv5s detection model based on deep learning, which increased the three-scale prediction of the model to four-scale prediction and improved the detection ability of the model for small obstacles. Ye et al. [14] proposed an adaptive focusing target feature fusion network based on YOLOX to balance detection speed and detection accuracy for complex underground environments. Based on the MobileNetV3 large module structure, Gao et al. [15] adopted the CBAM attention mechanism module, which can better improve the network's feature-extraction ability of complex pixel information in the regions of interest in coal and gangue images. The above studies have proposed target-recognition methods based on deep learning, but these methods are biased towards coal and gangue recognition, anchor rods in belt foreign bodies, and roadway anchor-hole position detection rather than the actual roadway-supported anchor rods.

After accurately identifying the anchor rods, the positioning of supported anchor rods is the crucial problem to realize the positioning of the drill-anchor robot. Suppose the position information of the roadway anchor rods can be detected in real-time and accurately. In that case, it will make the drill-anchor robot complete the drilling and anchoring operation more efficiently and ensure production safety. At present, many scholars have conducted in-depth research on vision-based positioning technology. Ma et al. [16] proposed an improved census transform algorithm to obtain the disparity map of anchor rods and the depth information of the anchor image. They used the minimum circumscribed rectangle and the maximum circumscribed rectangle algorithm to extract the pixel coordinates of the feature points of the bolt, which can accurately obtain the position information of the anchor rods. However, due to the complexity of the calculation process, the running time is too long. Han et al. [17] used a dynamic template-matching algorithm combined with an RGB-D camera for positioning, which had good positioning accuracy under static and dynamic conditions. However, this method detected a single target and could not meet the requirements of simultaneous positioning of multiple anchor rods in coal-mine roadways. Wang et al. [18] improved the YOLOv5s model, added the SPD-Conv and collaborative attention mechanism, and restored the depth map to realize anchor-hole positioning based on a deep-learning model and depth camera. However, this positioning method

relies entirely on vision and deep-learning recognition, and positioning is vulnerable to environment and equipment. Cheng et al. [19] proposed a binocular vision measurement technique using four light spots as feature points for the position of a boring machine under complex backgrounds such as mixed illumination and low illumination. However, the final positioning error of this method is larger than that of other methods, and the positioning accuracy needs to be improved.

In summary, we focus on exploring a coal-mine roadway anchor-rod recognition and positioning method based on the fusion of image enhancement and multiattention mechanism to improve the YOLOv7 (the seventh version of the You Only Look Once) model (YOLOv7 + CLAHE + BiFormer + CBAM fusion). This method preprocesses the image obtained by the visual sensor to solve the problems of the traditional way in the low illumination and uneven lighting environment of the coal-mine roadway. By using the deep-learning algorithm, combining the depth image with the RGB image alignment and the least squares linear fitting method, the roadway anchor rod's automatic identification and precise positioning are realized. This method has high efficiency and accuracy and can adapt to the characteristics of roadway anchor rods in different mining areas, which has a strong practicability and popularization value.

The structure of the rest of this article is as follows. Section 2 gives the principle and steps of the identification and positioning method. Section 3 mainly introduces the structure and composition of the image enhancement and multiattention mechanism fusion-improved YOLOv7 model proposed in this paper. Section 4 presents the positioning method of the supported anchor rod, including sensor calibration, coordinate transformation, and data processing. The experimental deployment and results are given in Section 5, and the work of this paper is summarized in Section 6.

2. Method and Principle

Aiming at the problem of anchor-rod identification and positioning in coal-mine tunneling, we constructed a YOLOv7 + CLAHE + BiFormer + CBAM fusion model and proposed a bolt identification and positioning method based on this fusion model. The principle of identification and positioning is shown in Figure 1.

The complex environment of the coal-mine roadway has a significant influence on anchor-rod recognition and positioning. Under the influence of complex environmental factors, the YOLOv7 model has no noticeable improvement effect on image processing. To effectively improve the accuracy of anchor-rod recognition, the contrast-limited adaptive histogram equalization algorithm (CLAHE) is used to enhance the original data and the image quality. Combined with the BiFormer and the Convolutional Block Attention Module (CBAM) to strengthen the ability of feature extraction, the anchor-rod target-detection and recognition model of YOLOv7 + CLAHE + BiFormer + CBAM fusion is constructed. The positioning method of the supported anchor rod based on the target anchor-rod detection and recognition model is established. The method includes the supported anchor-rod detection and recognition model, the perception system calibration and target-solving model, target coordinate transformation, and the positioning information solution model.

(1) For the supported anchor-rod detection and recognition model, collect anchor sample data in the roadway environment by simulating the roadway and combining with the image data of the actual coal-mine roadway to establish a supported anchor-rod dataset; an improved YOLOv7 target-detection model is built to improve the detection performance of the model. The obtained anchor-rod dataset is used to train the visual recognition model, and the camera is used to detect the real-time image. The detection box and confidence of the target are identified, and the detection box is used as the region of interest (ROI) in the image;

(2) For perception system calibration and the target-solving model, the camera is calibrated to obtain the internal and external parameters of the camera. The depth map and RGB image of the roadway anchor rod are captured, and the corresponding ROI in the depth image is processed by median filtering and bilateral filtering to restore the practical

information in the depth map. The center point of the target-detection frame is used as the positioning target point, and the depth map is aligned with the RGB image to obtain the depth value of the target point and the two-dimensional coordinates in the image;

(3) For target coordinate transformation and the positioning information solution model, the obtained depth value and the two-dimensional coordinate of the target point are transformed to obtain the three-dimensional coordinate information of the target point in the camera coordinate system. Aiming at the problem of large *X*-axis coordinate deviation, the coordinate value is corrected by least squares linear fitting to realize the visual positioning of the bolt.

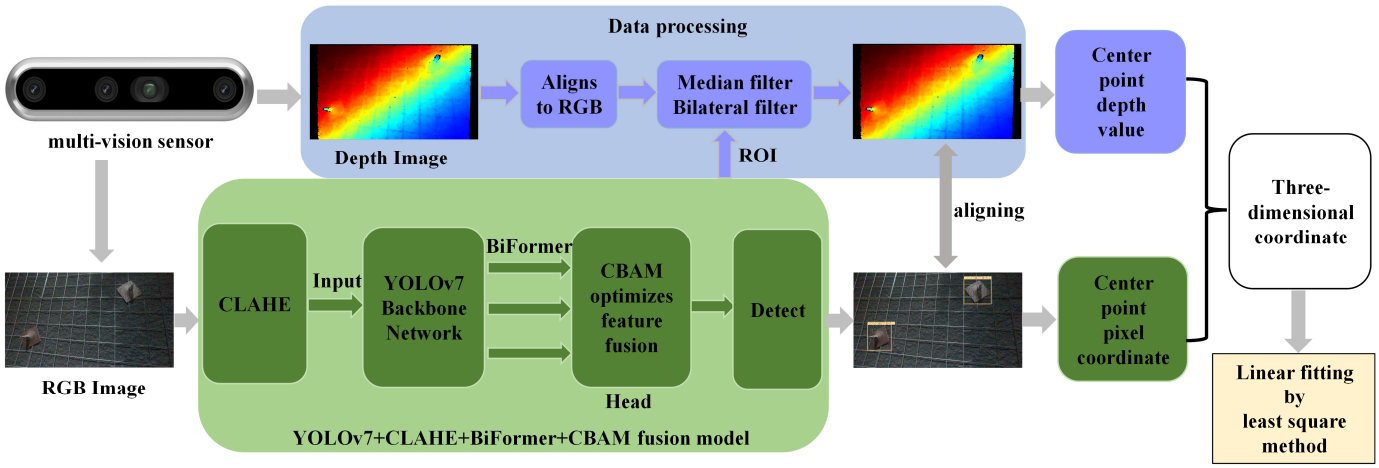


Figure 1. Principle of the anchor-rod identification and positioning method.

3. Construction of the Model

YOLOv7 [20–22] is a target-detection algorithm based on deep learning. It has real-time solid performance and can detect and locate multiple targets quickly and accurately in images. The main structure of the network includes the input layer, backbone, and head. Aiming at the problem of low and uneven lighting in coal-mine roadways and the difficulty of anchor-rod recognition caused by small targets, we propose to use the YOLOv7 target-detection algorithm in this paper, combined with CLAHE, BiFormer attention mechanism, and CBAM, to optimize the network structure and detection performance and, then, to improve the recognition accuracy and robustness of the model. The network structure is shown in Figure 2. The main model structure is as follows.

(1) Based on the YOLOv7 target-detection algorithm, the recognition and detection of small targets can be realized while ensuring the detection speed, and the center point position and target category of the detection box can be obtained;

(2) CLAHE enhances the images collected in low and uneven illumination environments, and precise sample data are obtained;

(3) The BiFormer attention mechanism is added at the end of the backbone network to improve the detection ability of small targets further;

(4) In the feature fusion stage, the CBAM combines channel and spatial attention so that the optimized feature map pays more attention to channel and spatial domain features, thus suppressing nonimportant information.

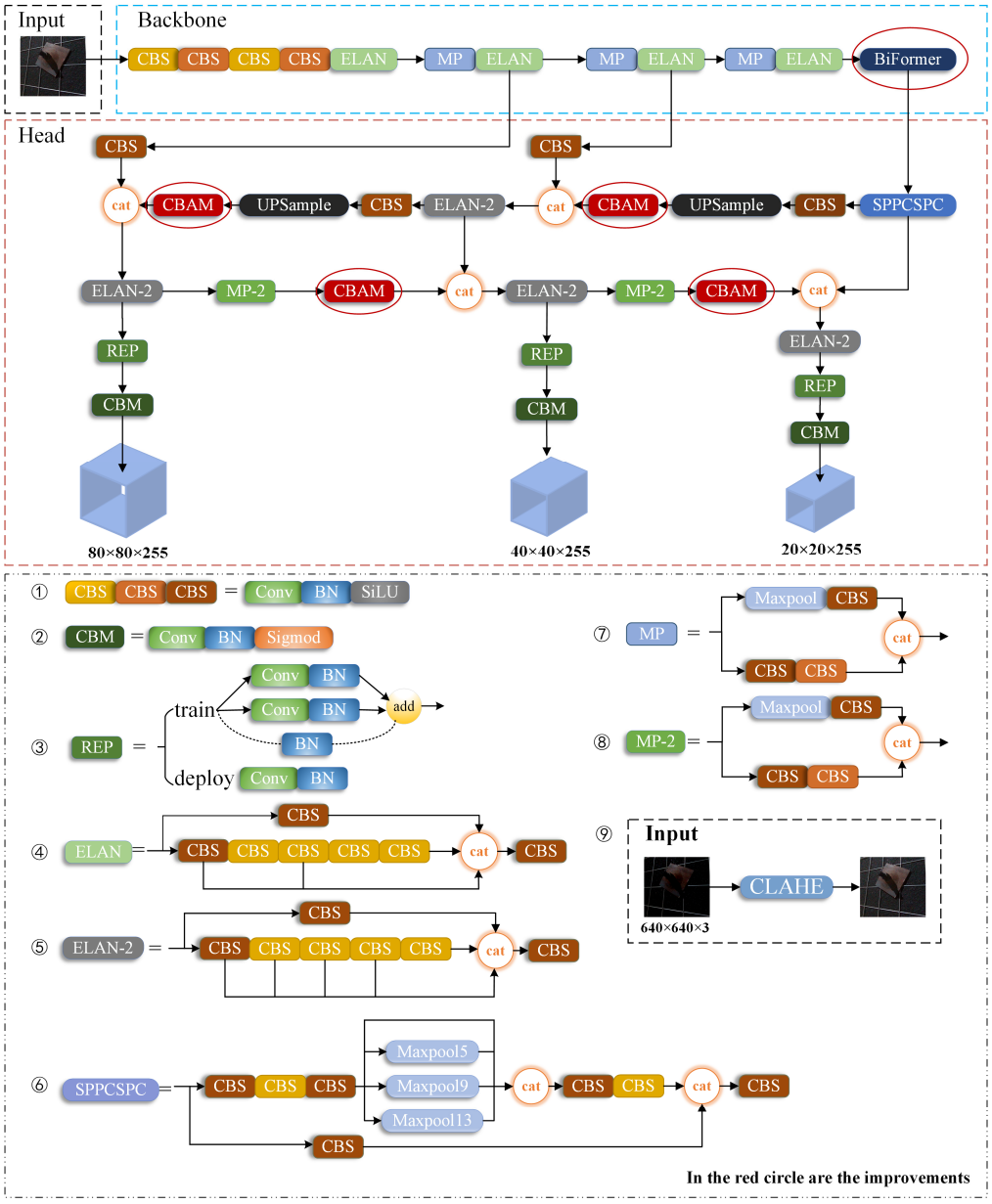


Figure 2. YOLOv7 + CLAHE + BiFormer + CBAM fusion model structure.

3.1. Image-Enhancement Processing

The preprocessing of the data by the YOLOv7 input layer cannot effectively improve the problem of unclear samples under low and uneven illumination conditions. Therefore, the CLAHE [23] method optimizes the images. The traditional histogram equalization algorithm (HE) is only for the global. The processing effect is not ideal when the image is locally too bright or too dark, and the histogram equalization will enhance the background noise. To solve the above two problems, the image clarity is enhanced by the Contrast-Limited Adaptive Histogram Equalization algorithm on the image subregions. Figure 3 is the flow chart of the CLAHE algorithm. In the collected original image, the anchor rods are in a dark environment, so the targets are unclear, and accurate identification and positioning cannot be achieved. Using the CLAHE method to enhance the image, the contrast limitation threshold is set to three, and the subregion size is (8,8). The contours of the anchor rods can be seen. The image clarity and brightness are better than the original image, which can be used as the input for target detection and guarantee subsequent recognition and positioning.

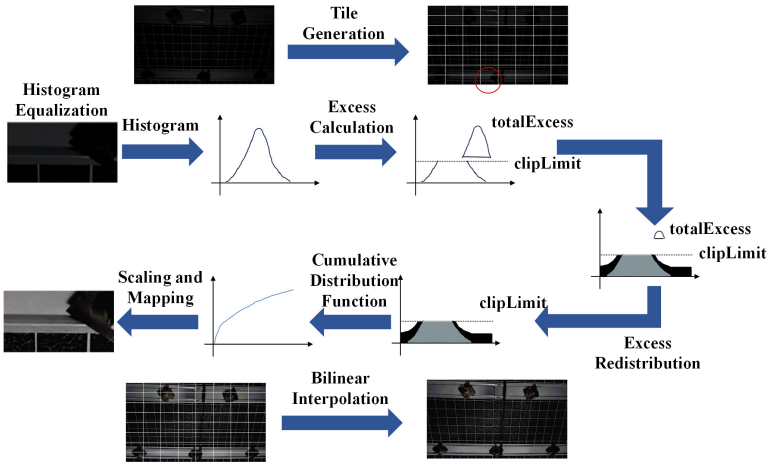


Figure 3. CLAHE algorithm flow chart. The red circle in the figure shows the part before and after the algorithm processing as a comparison.

3.2. Feature-Extraction Optimization Based on BiFormer

Aiming at the problems of multiple anchors and small samples in a single frame image, the sample contour and features are easily ignored under the influence of the environment, resulting in poor recognition results. The attention mechanism is added to the YOLOv7 target-detection network to strengthen the critical feature information. For the backbone, the BiFormer attention mechanism is added to the top layer, that is, the final extraction feature layer, to make up for the deficiency of the original model in small target-detection ability. The structure is shown in Figure 4. By embedding BiFormer's Bi-Level Routing Attention (BRA) module into YOLOv7, the vector of each position can interact with the vector of other places, and different weights are assigned according to the relative importance between them [24,25]. In BRA, each encoder contains multiple Transformer [26,27] attention heads. Through parallel computing, it captures different scales and types of information, learns the correlation between other image regions, pays attention to the union between areas, and uses sparsity to skip the most irrelevant computing regions, further reducing the computational pressure and difficulty.

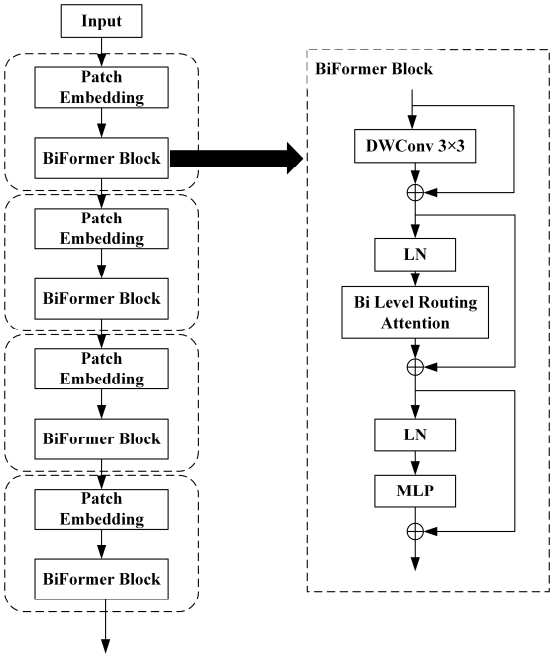


Figure 4. BiFormer attention mechanism structure.

3.3. Feature Fusion Optimization Based on CBAM

Based on strengthening the ability to detect small targets, extract effective features further, reduce the computational pressure, and improve the recognition accuracy, the CBAM is added after the head layer upsampled the results and the MP-2 structure. The CBAM structure is shown in Figure 5. Entering the feature map F, $F \in R^{C\times H\times W}$, and using the channel attention module to generate a one-dimensional channel attention map M_C,

$$\begin{aligned} M_C(F) &= \sigma(MLP(AvgPool(F)) + MLP(MaxPool(F))) \\ &= \sigma(W_1(W_0(F^C_{avg})) + W_1(W_0(F^C_{\max}))) \end{aligned} \tag{1}$$

where σ is the Sigmoid function, $W_0 \in R^{C/r\times C}$, $W_1 \in R^{C\times C/r}$, F^C_{avg}, and $F^C_{\max}$ represent the average pooling feature and the maximum pooling feature, respectively. The attention output of the channel is F', and the formula is:

$$F' = M_C(F) \otimes F \tag{2}$$

The spatial attention module performs average pooling and maximum pooling operations along the channel axis [28] and then uses a standard convolutional layer to obtain a two-dimensional spatial attention map M_S,

$$\begin{aligned} M_S(F) &= \sigma(f^{7\times7}([AvgPool(F); MaxPool(F)])) \\ &= \sigma(f^{7\times7}([F^S_{avg}; F^S_{\max}])) \end{aligned} \tag{3}$$

where σ is the Sigmoid function, and $f^{7\times7}$ is the convolution kernel of 7×7. The spatial attention output is F'', and the formula is:

$$F'' = M_S(F') \otimes F' \tag{4}$$

As a lightweight module, CBAM can improve the network representation classification and detection performance by focusing on essential features [29].

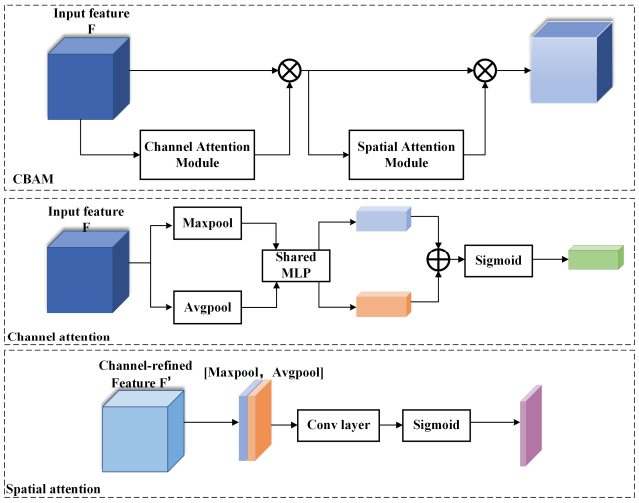


Figure 5. CBAM structure.

3.4. Construction of Mixed Dataset of Supported Anchor Rod

To improve the safety and production efficiency of coal-mine roadways, it is crucial to accurately identify and detect anchor rods in tunnels. The target-detection and recognition model requires a large amount of sample data during training. However, because there are few studies on the recognition of supported anchor rods in coal-mine roadways, and it is not easy to obtain data in coal mines; visual sensors are mainly used to obtain anchor-rod images in simulated roadways and establish the sample dataset. To improve the recognition

accuracy of the supported anchor rods in the uneven illumination environment, based on the data collected in the simulated roadway environment, some anchor-rod images in the natural underground roadway environment are added to construct a mixed dataset of anchor-rod images in the laboratory environment–real environment. The dataset contains a total of 2958 images. Part of the supported anchor-rod dataset is shown in Figure 6. To expand the number of samples in the dataset, data-processing methods, such as random rotation, image brightness enhancement, contrast enhancement, and image stitching, were used to form a total of 5174 sample datasets of supported anchor rods, which are divided into training, testing, and validation sets according to the ratio of 7:2:1.



Figure 6. Part of the supported anchor-rod dataset.

4. Positioning Method of Supported Anchor Rod in Roadway

4.1. MultiVision Sensor Calibration

To align the depth map with the RGB image and accurately locate the anchor rod, it is necessary to determine the internal parameters of the sensor and the rotation and translation relationship between the camera coordinate systems of the multivision sensor. Because the multiview vision sensor can capture both the depth map and the RGB image simultaneously, we do not need to consider the timestamp synchronization problem of each camera in this paper and only consider the spatial conversion relationship. The RGB monocular camera obtains the RGB image on the sensor, and the binocular camera receives the depth image on the left and right sides. The multivision sensor is shown in Figure 7.

The calibration of the multivision sensor is to solve the transformation relationship between the coordinate systems and the distortion parameters of each camera. The internal and external parameters and distortion parameters of the sensor are calibrated by the Zhang Zhengyou calibration method [30]. The calibration results are shown in Table 1.

Table 1. Sensor calibration parameters.

Parameter	Left Imager	Right Imager	RGB
Focal length	$[381.5116 \quad 383.3934]$	$[381.2979 \quad 383.5978]$	$[378.7371 \quad 380.0363]$
Intrinsic matrix	$\begin{bmatrix} 381.5116 & 0 & 312.2119 \\ 0 & 383.3934 & 239.3964 \\ 0 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 381.2979 & 0 & 312.4911 \\ 0 & 383.5978 & 239.1390 \\ 0 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 378.7371 & 0 & 320.3359 \\ 0 & 380.0363 & 241.0257 \\ 0 & 0 & 1 \end{bmatrix}$
Radial distortion	$[0.0018 \quad -0.0032]$	$[0.0028 \quad -0.0033]$	$[-0.0492 \quad 0.0479]$
Tangential distortion	$[-6.3533e^{-4} \quad -0.0012]$	$[9.4072e^{-4} \quad -0.0016]$	$[-0.0012 \quad 2.7327e^{-4}]$
Reprojection error	$[3.5172e^{-5} \quad 5.8203e^{-5}]$	$[2.1552e^{-5} \quad 5.5460e^{-5}]$	$[3.3221e^{-5} \quad 6.0154e^{-5}]$
Rotation matrix		$\begin{bmatrix} 1.0000 & 0.0005 & -0.0001 \\ -0.0005 & 1.0000 & 0.0010 \\ 0.0001 & -0.0010 & 1.0000 \end{bmatrix}$	$\begin{bmatrix} 1.0000 & -0.0005 & 0.0013 \\ 0.0005 & 1.0000 & 0.0001 \\ -0.0013 & -0.0001 & 1.0000 \end{bmatrix}$
Translation matrix		$[-0.0951 \quad 6.6814 \quad 0.0006]$	$[0.0361 \quad -4.4994 \quad -0.0004]$

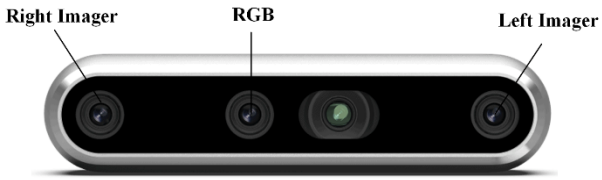


Figure 7. Multivision sensor.

4.2. Coordinate Transformation Model

In the positioning process of the supported bolt, to determine the corresponding relationship between the three-dimensional coordinates of the target point of the supported anchor rod and its image at a particular time, it is necessary to establish a camera-pixel coordinate system conversion model. According to the position of the target point in the pixel coordinate system, the position coordinates relative to the RGB camera are obtained by the conversion model. The conversion relationship between the camera coordinate system and the pixel coordinate system is shown in Figure 8. In the figure, P is the real target point, P' is the corresponding point of the target point on the image, uv is the pixel coordinate system, xoy is the image coordinate system, and O_L, O, O_R are the origin of the camera coordinate system of the left camera, RGB camera, and right camera, respectively.

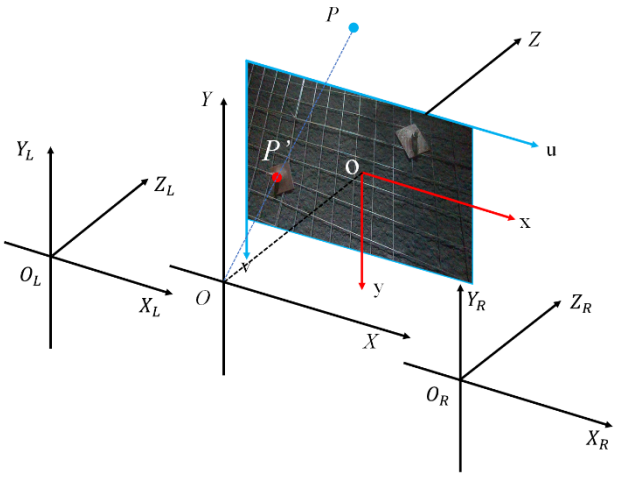


Figure 8. Coordinate transformation relationship.

The image and pixel coordinate systems are on the same imaging planes. The measurement unit and the origin position of the two are different. The coordinate transformation formula is:

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{dx} & 0 & u_0 \\ 0 & \frac{1}{dy} & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \tag{5}$$

where (u, v) is the coordinate of the target point in the pixel coordinate system, and the unit is the pixel; (x, y) is the coordinate of the target point in the image coordinate system, the unit is mm; dx and dy represent the actual size of a single pixel along the x and y axes, and the unit is mm; and (u_0, v_0) is the coordinate value of the origin of the image coordinate system in the pixel coordinate system.

The conversion relationship formula between the camera coordinate system and the image coordinate system is:

$$Z \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \tag{6}$$

where f is the focal length of the camera and (X, Y, Z) is the coordinates of the target point in the camera coordinate system.

Combining Formulas (5) and (6), the conversion relationship between the camera coordinate system and the pixel coordinate system can be obtained as follows.

$$Z \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{dx} & 0 & u_0 \\ 0 & \frac{1}{dy} & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \tag{7}$$

To obtain the depth value of the target point, we choose to capture the depth map and RGB image simultaneously. Because the RGB camera responsible for capturing the RGB image in the multiview vision sensor and the left and right infrared cameras accountable for capturing the depth map have different installation positions and internal parameters, the pixels on the image do not match each other, and aligning the RGB image to the depth map will distort the scene in the color image. In contrast, the pixels in the depth map only contain depth information. Therefore, the depth map can be aligned with the RGB image to ensure that the same pixel position in the two images corresponds to the same target point. Figure 9 shows the effect of the depth map aligned to the RGB image. The left side is the original depth map captured by the left and right infrared cameras, and the right side is the depth map after alignment. The same pixel area in the depth map before and after alignment is extracted, and the position of the bolt can be seen to change significantly. The alignment process needs to convert the coordinates between the three lenses according to the rotation and translation relationship between the cameras in Table 1 to merge the three independent visual channels. The conversion process is as follows.

According to the above coordinate transformation relationship, the transformation relationship of the left infrared camera from the camera coordinate system (X_L, Y_L, Z_L) to the pixel coordinate system (u_L, v_L) is as follows.

$$Z_L \begin{bmatrix} u_L \\ v_L \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{dx_L} & 0 & u_{L0} \\ 0 & \frac{1}{dy_L} & v_{L0} \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} f_L & 0 & 0 & 0 \\ 0 & f_L & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X_L \\ Y_L \\ Z_L \\ 1 \end{bmatrix} \tag{8}$$

The conversion formula from the left infrared camera coordinate system to the RGB camera coordinate system is:

$$\begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} = M * \begin{bmatrix} X_L \\ Y_L \\ Z_L \\ 1 \end{bmatrix} \tag{9}$$

where the conversion matrix M is obtained by camera calibration, and the conversion relationship between the pixel coordinate system of the left infrared camera and the pixel coordinate system of the RGB camera is obtained by combining Formulas (8) and (9). The same is true for the right infrared camera. The coordinate transformation is completed, and the alignment of the depth and RGB images can be realized.

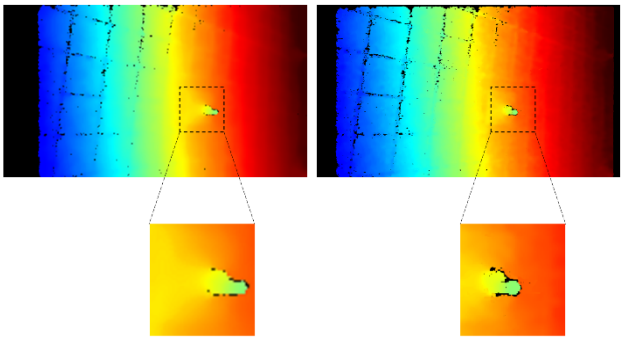


Figure 9. Depth map alignment effect.

4.3. Supported Anchor-Rod Data Processing

The positioning of the supported anchor rod must ensure that the depth map contains effective information of the target point to obtain the depth value. However, in collecting the depth image, some information is in the depth.

The image will inevitably be missing due to noise interference. As shown in Figure 10, the black area in the picture is the information loss caused by the depth map noise. Therefore, the depth image needs to be denoised after the depth map is aligned with the RGB image.

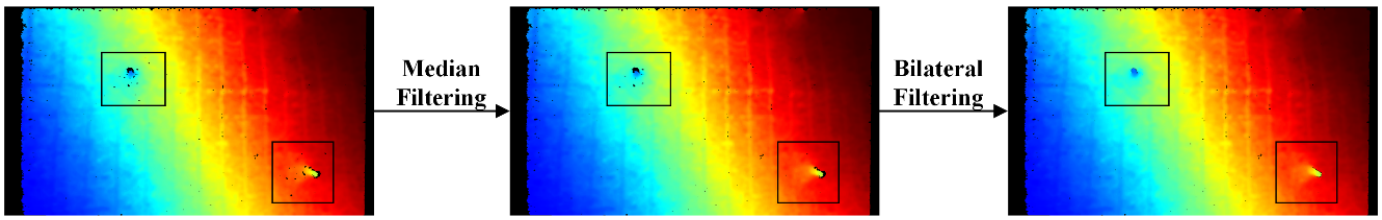


Figure 10. Depth map denoising effect. The black box in the figure is ROI.

Extracting a two-dimensional region of interest (ROI) combined with a median and bilateral filtering method is used to denoise the depth map of the supported anchor rod. The main process of the data-processing method is as follows.

(1) The improved YOLOv7 model recognizes the RGB image to obtain the anchor-rod detection boxes. The detection boxes are used as the ROI, and the corresponding area of the aligned depth map is extracted to reduce the unnecessary calculation;

(2) The median filtering method is used to denoise the image region in ROI. The median filtering traverses the pixel values of each pixel in the area and its adjacent pixels. After sorting, the pixel value in the middle position is used as the pixel value of the current pixel to eliminate noise [31]. Because the depth value of each pixel in the ROI of the depth map is similar, and the depth value of the noise point is quite different, the noise is filtered

by the median filtering method. As shown in Figure 10, the black area in the image is significantly reduced, but there is still interference;

(3) Therefore, bilateral filtering is used for further denoising to retain the edge information of the image better while filtering out the noise. The bilateral filtering formula can be expressed as:

$$I_{filtered}(x,y) = \frac{1}{W_P(x,y)} \sum_{(i,j)\in\Omega} \begin{matrix} I(x+i,y+i)\cdot F_{spatial}(i,j) \\ \cdot F_{range}(I(x,y),I(x+i,y+i)) \end{matrix} \quad (10)$$

where $I_{filtered}(x,y)$ is the pixel value after filtering, $I(x,y)$ is the pixel value in the original image, $W_P(x,y)$ is the normalized weight, Ω is the neighborhood window of the filter, $F_{spatial}(i,j)$ is the spatial domain kernel function, and $F_{range}I(x,y)$ is the pixel value domain kernel function. As shown in Figure 10, the black area disappears completely, and the target point can be further extracted for anchor-rod positioning.

Through the YOLOv7 model, the size and position coordinates of the detection frame are obtained, and the center point of the anchor detection frame is used as the target point to realize the positioning of the anchor rod. When the drill-anchor robot moves forward along the roadway, the multivision sensor is located at the top of the fuselage to shoot the side plate, and the depth image is used to calculate the depth information. The installation position of the multivision sensor is known, and the anchor-rod row spacing and bolt spacing in each row of anchor rods are relatively fixed. However, due to the nonuniform speed of the drill-anchor robot and the calculation error in the coordinate conversion process, we use the least square method to perform linear fitting on each row of anchor rods on the side plate to reduce the positioning error in the *X*-axis direction of the camera coordinate system and correct the coordinate value. The positioning model is shown in Figure 11. In the figure, L is the distance between the multivision sensor and the sidewall plate. For the side-plate anchor rods, the coordinates in the *Z*-axis direction can be ignored, and only the coordinate information of the X and Y axes can be linearly fitted.

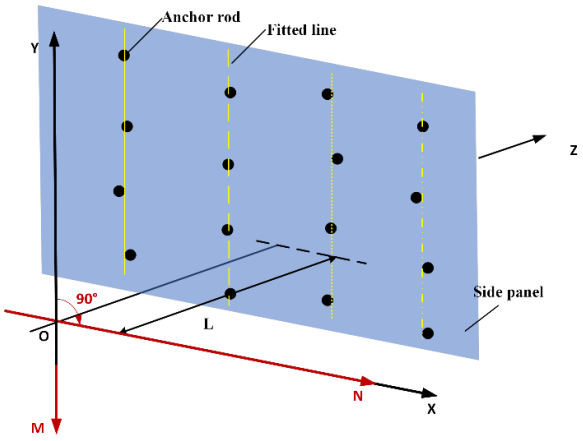


Figure 11. Positioning model diagram.

The fitting line is required to be perpendicular to the *X*-axis to obtain the corrected coordinates. However, the least square method is unsuitable for fitting a straight line perpendicular to the *X*-axis. Therefore, the *XOY* coordinate system is rotated 90° clockwise to obtain a new plane coordinate system *MON*. The Y coordinate value of the target point is reversed, and the X and Y coordinate values are adjusted to obtain the coordinate value (*m*,*n*) of the target in the new coordinate system. Then, the straight-line fitting is carried out, and the slope of the fitted straight line is required to be zero. The least squares linear fitting formula is as follows.

$$N = aM + b \quad (11)$$

$$a = \frac{Q\sum(m_i n_i) - \sum m_i \sum n_i}{n\sum {m_i}^2 - \left(\sum m_i\right)^2} \tag{12}$$

$$b = \frac{\sum m_i^2 \sum n_i - \sum m_i \sum (m_i n_i)}{Q\sum m_i^2 - \left(\sum m_i\right)^2} \tag{13}$$

where Q is the number of fitting points and a is the slope of the fitted straight line in the *MON* coordinate system. In this paper, $a = 0$ should be set; b is the N-axis intercept.

The positioning method is as follows.

(1) In the same frame image, the position of the detection boxes of multiple targets is obtained, and the pixel coordinates of the center point of the detection box are calculated;

(2) The depth image is aligned with the RGB image before denoising, and the depth value Z of the target point under the same pixel coordinate is obtained;

(3) The coordinate value of the target point is converted from the pixel coordinate system to the RGB camera coordinate system, and the position of the target point relative to the camera is obtained. The anchor-rod coordinate is fitted by the least square method to correct the error, and the final position coordinate of the anchor rod is obtained to realize the positioning of the supported anchor rod.

5. Experiment

5.1. Experiment Deployment

As shown in Figure 12, it is a coal-mine roadway simulation experiment environment to verify the feasibility of the anchor-rod identification and positioning method. Based on the multivision sensor to collect the roadway environment, as shown in Table 2, the main parameters of the visual sensor are constructed, and the experimental detection platform shown in Figure 12 is constructed. The simulation of the bolt row spacing in the roadway is 0.8 m, the spacing of the single row of anchor rods is 0.8 m, and the multivision sensor is 1.5 m away from the side plate. The experimental principle and process are shown in Figure 13.

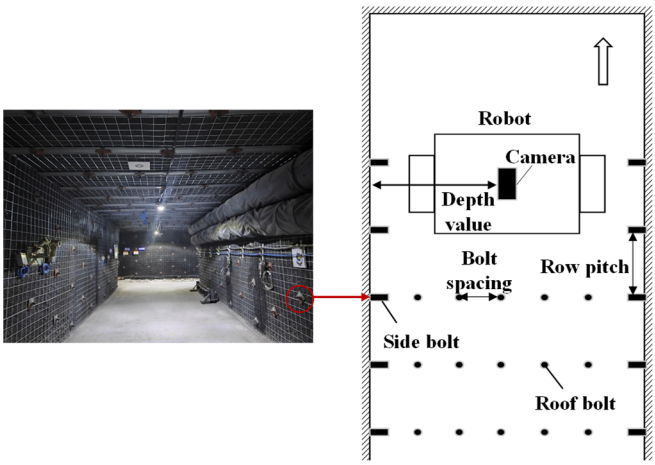


Figure 12. Coal-mine roadway simulation experiment environment and platform.

Table 2. Main parameters of the sensor.

Parameter	Value
Ideal range	0.6–6 m
Length × Depth × Height	124 mm × 26 mm × 29 mm
Field of view	86° ± 3° × 57° ± 3°
Depth output resolution	1280 × 720
RGB frame resolution	1280 × 720
Frame rate	30 fps

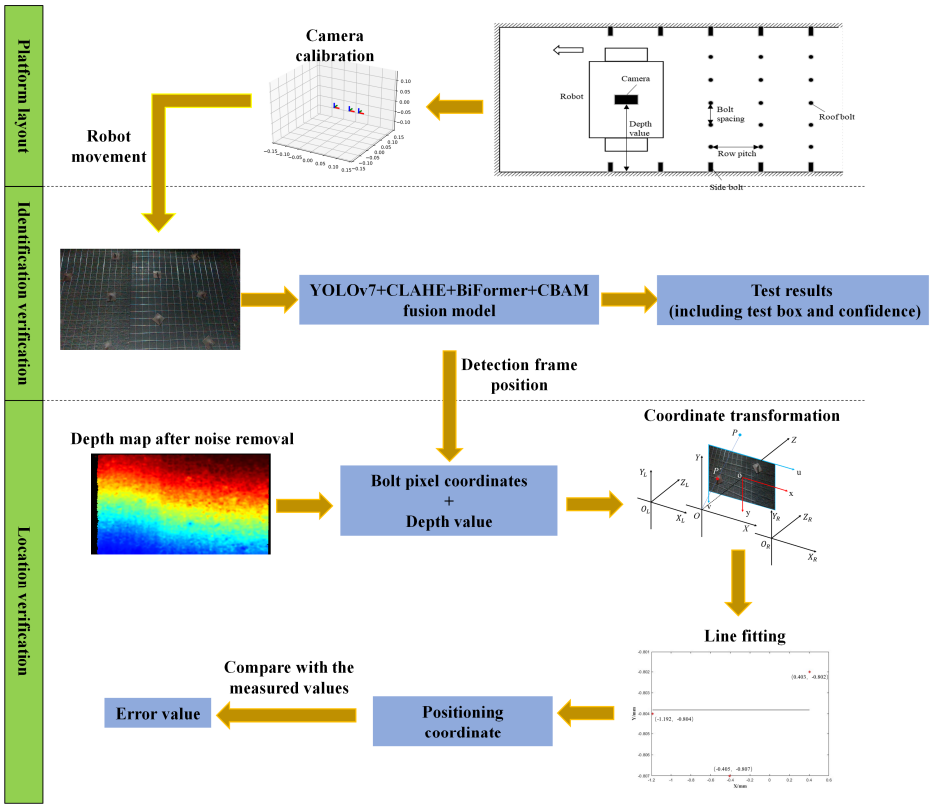


Figure 13. Experimental principle and flow chart.

Before the experiment, the internal and external parameters of the sensor were determined by calibration, and the RGB camera optical center was used as the coordinate origin. The direction of the roadway parallel to the side panel is the x-axis, the direction perpendicular to the roof is the y-axis, and the direction perpendicular to the side panel is the z-axis. The detection model proposed in this paper is used to identify and detect the anchor rod, and the model's performance is verified by ablation experiments. The pixel coordinates of the target point are obtained through the model, the corresponding depth values are obtained in the depth map, and the positioning results are converted into coordinates in the RGB camera coordinate system. The coordinate value is corrected by linear fitting to obtain the position information of the target point. Because there will be invalid areas in the depth map when the distance is too far, only eight anchors in three rows are located in the experiment. By comparing the measured value with the actual value and the fitting value, the accurate positioning of the anchor rod is finally realized.

5.2. Experiment and Analysis of Target-Detection and Recognition Model

The improved YOLOv7 model can effectively improve the quality of input images and reduce the influence of illumination factors on recognition. Adding an attention mechanism can improve the detection effect of small targets and the attention to practical features, reduce the computational pressure, and improve the recognition accuracy.

To verify the actual detection effect of the proposed improved YOLOv7 model, the ablation experiment under the same conditions was carried out using the self-built dataset. The model evaluation indexes include precision (P), recall (R), mean average precision (mAP), floating-point operations per second (FLOPs), parameters, and the detection speed (FPS). The calculation formulas of P, R, and mAP are as follows.

$$P = \frac{TP}{TP + FP} \tag{14}$$

$$R = \frac{TP}{TP + FN} \tag{15}$$

$$mAP = \frac{\sum_{i}^{N} AP_i}{N} \tag{16}$$

where TP is the number of correct anchor rods detected, FP is the number of other objects seen as anchor rods by mistake, FN is the number of the right anchor rods detected as other objects by mistake, N is the number of detection categories, and AP is the average accuracy of a single category. The IOU threshold is set to 0.5; that is, the average accuracy of the single category is $AP_{0.5}$, and the detection category N is one, so mAP = $AP_{0.5}$. The results of ablation experiments are shown in Table 3.

Table 3. Comparison of ablation experiment results.

Subject	Model	Precision	Recall	mAP	FLOPs/G	Parameters/M	FPS
a	YOLOv7	90.2%	90.1%	89.2%	103.5	37	87.5
b	CLANE + YOLOv7	92.6%	93.8%	91.0%	103.5	37	87.5
c	CLAHE + CBAM + YOLOv7	94.4%	93.5%	92.7%	102.1	35	128.3
d	CLAHE + BiFormer + YOLOv7	94.0%	91.2%	92.1%	103.7	39	122.6
e	CLAHE + BiFormer + CBAM + YOLOv7	96.2%	94.7%	94.9%	105.8	35	123.2

The ablation experiment results show that model a is the 0.1 version of the original YOLOv7 model; model b adds the CLAHE algorithm to enhance the data based on model a, which reduces the influence of illumination conditions on the image. The *mAP* is 1.8% higher than model a, indicating that CLAHE can significantly improve image clarity, thus improving the accuracy of anchor-rod recognition. Model c adds the CBAM based on model b. The number of parameters and FLOPs are reduced, and the *mAP* is 3.5% higher than the original model. It is proven that the introduction of the CBAM makes the model pay more attention to practical features and filter out invalid information, which is of practical significance for improving the model's performance. Model d adds the BiFormer attention mechanism based on model b. Compared with the original model, *mAP* is increased by 2.9%, and the number of parameters is also increased. This confirms that the BiFormer attention mechanism helps improve the detection ability of small targets but simultaneously increases the amount of calculation. Based on model d, the CBAM is added to the head network of model e, and the *mAP* is 5.7% higher than that of the original model. At the same time, the increase in the number of parameters caused by the BiFormer attention mechanism is suppressed to a certain extent. The comparison of the detection effects of the original YOLOv7 model and the improved model is shown in Figure 14.

The experimental results show that the target-detection model based on YOLOv7 + CLAHE + BiFormer + CBAM fusion has a good effect on anchor-rod recognition in low and uneven illumination environments. The *mAP* is 5.7% higher than that of the YOLOv7 model, and the number of parameters is reduced by 2M, which has a significant improvement in detection accuracy.

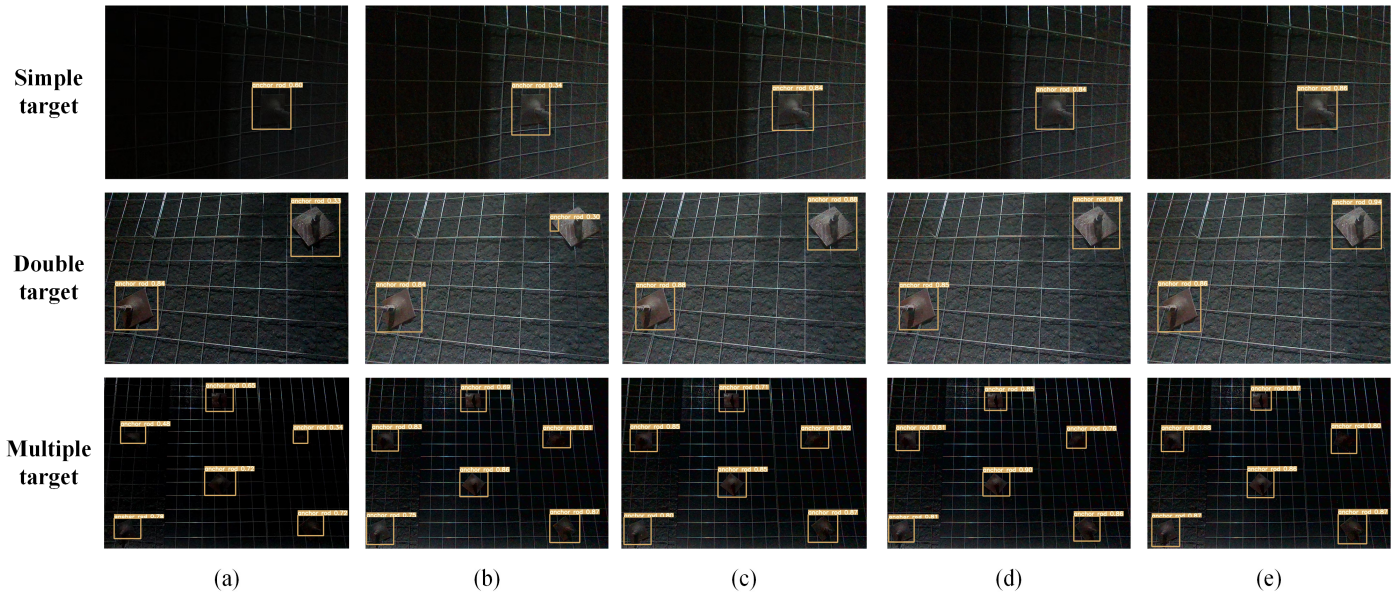


Figure 14. Comparison of experimental results. (**a**) YOLOv7. (**b**) CLAHE + YOLOv7. (**c**) CLAHE + CBAM + YOLOv7. (**d**) CLAHE + BiFormer + YOLOv7. (**e**) CLAHE + BiFormer + CBAM + YOLOv7.

5.3. Experimental Verification of Anchor-Rod Positioning

Figure 15a is the result of the improved YOLOv7 model. The samples in the picture are error free, and the confidence is higher than the threshold. At the same time, the size and location information of each detection box are obtained to calculate the pixel coordinates of the center point. Figure 15b is the corresponding depth map after alignment. The edge of the depth map is invalid information, and the information of the middle three rows of anchor rods is complete. The detection frames extracted from the RGB image are mapped to the depth map as ROI, and only the depth map in ROI is denoised to complete the depth information and extract the depth value of the center point of the detection frame.

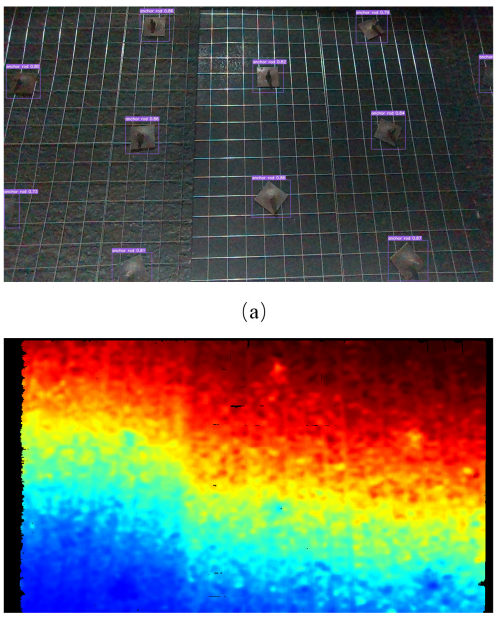


Figure 15. (**a**) Network detection effect. (**b**) Depth rendering after denoising the positioning experimental verification.

Through the coordinate transformation of the pixel coordinates, combined with the depth value, the position coordinates in the RGB camera coordinate system are obtained, and the three rows of anchor rods are fitted with straight lines to get the corrected position coordinates. For example, Figure 16 shows the fitting lines for the three rows of anchor rods. The pixel coordinate value, depth value, coordinate measurement value, conversion coordinate value, and fitting coordinate value of the target point are shown in Table 4.

Table 4. Comparison of positioning.

Row Number	Pixel Coordinate Value/Pixel	Depth/m	Coordinate Measurement Value/m	Conversion Coordinate Values/m	Fitting Coordinate Value/m	Absolute Value of an Error/mm
1	(384,53)	1.506	(−0.805,1.200,1.500)	(−0.804,1.192,1.506)	(−0.804,1.192,1.506)	(1,8,6)
	(361,341)	1.503	(−0.805,0.410,1.500)	(−0.807,0.405,1.503)	(−0.804,0.405,1.503)	(1,5,3)
	(330,680)	1.496	(−0.810,−0.400,1.500)	(−0.802,−0.403,1.496)	(−0.804,−0.403,1.496)	(6,3,4)
2	(690,186)	1.507	(0.112,0.800,1.500)	(0.126,0.799,1.507)	(0.114,0.799,1.507)	(2,1,7)
	(698,500)	1.502	(0.110,0.050,1.500)	(0.103,0.051,1.502)	(0.114,0.051,1.502)	(4,1,2)
3	(961,57)	1.508	(0.836,1.220,1.500)	(0.831,1.225,1.508)	(0.835,1.225,1.508)	(1,5,8)
	(1003,329)	1.502	(0.840,0.450,1.500)	(0.836,0.454,1.502)	(0.835,0.454,1.502)	(5,4,2)
	(1055,663)	1.498	(0.828,−0.450,1.500)	(0.839,−0.448,1.498)	(0.835,−0.448,1.498)	(7,2,2)

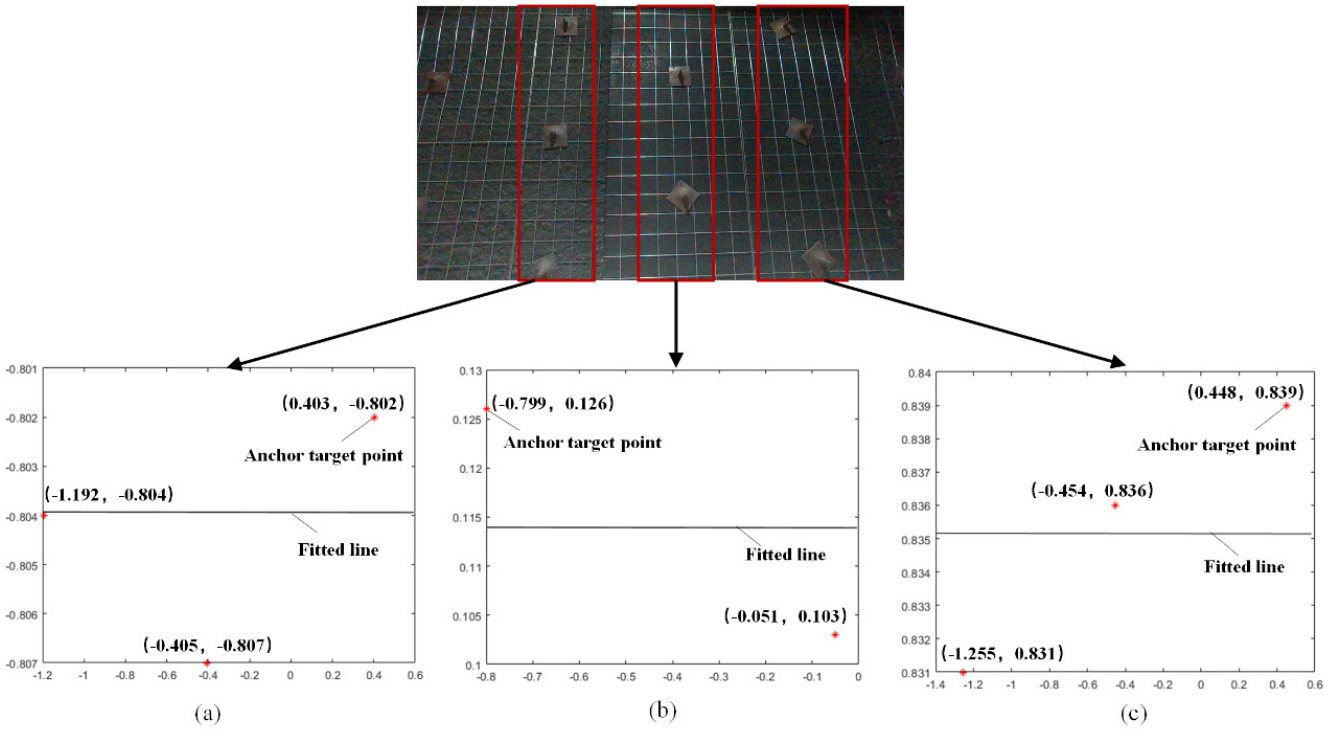


Figure 16. Fitting straight-line diagram. (**a**) The first row of anchor rods. (**b**) The second row of anchor rods. (**c**) The third row of anchor rods.

Through Figure 15, the target-detection model of the YOLOv7 + CLAHE + BiFormer + CBAM fusion is reliable for anchor-rod recognition in low and uneven illumination environments and can accurately detect targets. According to the data comparison in Table 4, it can be concluded that the depth value of the target point obtained by the depth map combined with the RGB image has a minor error with the actual measured value. The *Y*-axis coordinate has no significant deviation from the measured value, but the *X*-axis coordinate has a large error in individual cases. By fitting the straight line to correct the coordinates, the more significant *X*-axis coordinate deviation is improved.

Figure 17 shows the positioning error of 60 anchor rods in 15 rows of simulated roadway on the X, Y, and Z axes. The positioning error does not exceed 11 mm, proving that the positioning method in this paper can accurately locate the anchor-rod target in the image, effectively reducing the error caused by environment or coordinate transformation, and has good robustness.

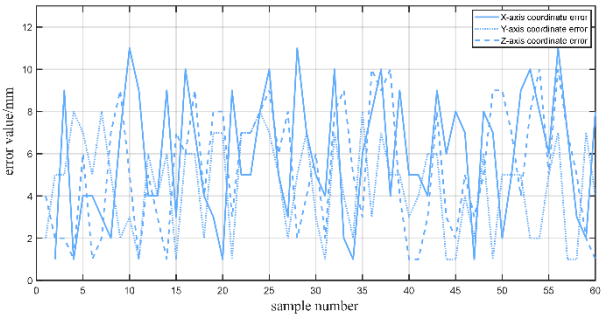


Figure 17. Coordinate error curve.

6. Conclusions

Aiming at the problems that anchor-rod recognition is difficult, and the positioning accuracy cannot be guaranteed due to the low illumination and uneven lighting in a coal-mine roadway, we propose an anchor identification and positioning method based on the fusion of image enhancement and multiattention mechanism that improved the YOLOv7 model. The process takes the roadway-supported anchor rods as the recognition and positioning target. It uses the improved YOLOv7 target-detection model to identify the anchor rods, which effectively solves the influence of underground illumination on the recognition accuracy of the supported anchor rods and improves the recognition accuracy. The method of depth map combined with least square linear fitting is used to locate the anchor rods, which provides an effective way for the accurate positioning of the anchor rods in the coal-mine roadway.

(1) The proposed image enhancement and multiattention mechanism fusion improved the YOLOv7 target-detection model, which can effectively reduce the difficulty of target recognition caused by illumination problems, enhance the ability to extract small target features, and improve recognition accuracy;

(2) A multiview vision sensor is used to collect sample images in simulated roadways and is combined with the anchor-rod data of actual coal-mine roadways, which enhances the diversity of the samples and the adaptability of the model to the supported anchor rods of the actual underground roadway and establishes the image dataset of the anchor-rod target detection under the environment of low and uneven illumination;

(3) The image is enhanced by the Contrast-Limited Adaptive Histogram Equalization algorithm, which reduces the influence of the illumination problem in anchor-rod recognition, improves the clarity of target contour and features, weakens noise interference, and improves image quality;

(4) Based on the YOLOv7 model, BiFormer and CBAM are added to improve the detection ability of the model for small targets and the attention to essential features and reduce the noise interference of the environment;

(5) The method of extracting ROI combined with median and bilateral filtering is used to denoise the depth map, which solves the problem of effective information loss caused by noise. At the same time, the processing time and calculation amount are reduced, which is beneficial to extract the depth value of the target point;

(6) Experiments verify the proposed recognition and positioning method of the supported anchor rod. The experimental results show that the mAP of the improved model is 5.7% higher than that of the original model, reaching 94.9%. The model can identify the supported anchor rod accurately. The error between the positioning coordinates of the target point on each axis and the measured coordinates is no more than 11 mm, which

meets the positioning-error requirements of the anchor rod in the drilling and anchoring operation. It shows that the positioning method can ensure the positioning accuracy of the anchor rod and provide a basis for the positioning of the body of the drill-anchor robot.

Author Contributions: Software, Validation, E.Z.; Investigation, C.W.; Data curation, X.Y., Y.Q.; Writing—original draft, J.Y.; Writing—review & editing, X.X.; Supervision, Q.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (Grant Nos. 52374161, 51834006); the National Key Research Development Program Young Scientists Project of China (No. 2022YFF0605300); Xi'an Science and Technology Plan Project (No. 22GXFW0067); National Key Research Development Program of China (No. 2023YFC2907603); and Shaanxi Province "Two Chain" Integrated Enterprise (Institute) Joint Project (No. 2023-LL-QY-03).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The supported anchor-rod dataset presented in this article is not publicly available due to the data being part of ongoing research.

Conflicts of Interest: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. Kang, H.P.; Jiang, P.F.; Liu, C.; Wang, Z.Y.; Luo, C.; Guo, J.C.; Chen, Z.L.; Cao, X.M. Current situation and development trend of rock bolting construction equipment in coal roadway. *J. Mine Autom.* **2023**, *49*, 1–18. [CrossRef]
2. Wang, B.K. Current status and trend analysis of roadway driving technology and equipment in coal mine. *Coal Sci. Technol.* **2020**, *48*, 1–11. [CrossRef]
3. Wang, H.; Chen, M.J.; Zhang, X.F. Twenty Years development and Prospect of Rapid Coal Mine Roadway Excavation in China. *J. China Coal Soc.* **2024**, 1–16. [CrossRef]
4. Wang, H.W.; Jin, L.; Yan, Z.R.; Guo, J.J.; Zhang, F.J.; Li, C. Roadway anchor hole recognition and positioning method based on image and point cloud fusion. *Coal Sci. Technol* **2024**, 1–15. Available online: http://kns.cnki.net/kcms/detail/11.2402.TD.20231114.1722.007.html (accessed on 13 January 2024).
5. Ge, S.R.; Hu, E.Y.; Li, Y.W. New progress and direction of robot technology in coal mine. *J. China Coal Soc.* **2023**, *48*, 54–73. [CrossRef]
6. Zhang, J. Research on motion control of manipulator of anchor drilling robot based on WOA-FOPID algorithm. *Coal Sci. Technol.* **2022**, *50*, 292–302. [CrossRef]
7. Ge, S.R.; Hu, E.Y.; Pei, W.L. Classification system and key technology of coal mine robot. *J. China Coal Soc.* **2020**, *45*, 455–463. [CrossRef]
8. Cheng, J.; Song, Z.L.; Li, H.; Ma, Y.Z.; Li, H.P.; Sun, D.Z. A novel image enhancement method via dual-branch coupled Transformer network for underground coalmine. *J. China Coal Soc.* **2024**, 1–12. [CrossRef]
9. Zhang, Y.; Tong, L.; Lai, X.P.; Cao, S.G.; Yan, B.X.; Liu, Y.Z.; Sun, H.Q.; Yang, Y.B.; He, W. Coal-rock interface perception and accurate recognition in tunneling space under coal dust environment based on machine vision. *J. China Coal Soc.* **2023**, 1–14. [CrossRef]
10. Zhang, F.J.; Wang, H.W.; Wang, H.R.; Li, Z.L.; Wang, Y.H. Intelligent identification and positioning of steel belt anchor hole in coal mine roadway support. *J. Mine Autom.* **2022**, *48*, 76–81. [CrossRef]
11. Hao, S.; Zhang, X.; Ma, X.; Sun, S.Y.; Wen, H.; Wang, J.L. Foreign object detection in coal mine conveyor belt based on CBAM-YOLOv5. *J. China Coal Soc.* **2022**, *47*, 4147–4156. [CrossRef]
12. Xu, J.-J.; Zhang, H.; Tang, C.-S.; Cheng, Q.; Tian, B.; Liu, B.; Shi, B. Automatic soil crack recognition under uneven illumination condition with the application of artificial intelligence. *Eng. Geol.* **2022**, *296*, 106495. [CrossRef]
13. Yang, T.; Wang, S.; Tong, J.; Wang, W. Accurate real-time obstacle detection of coal mine driverless electric locomotive based on ODEL-YOLOv5s. *Sci. Rep.* **2023**, *13*, 17441. [CrossRef]
14. Ye, T.; Zheng, Z.; Li, Y.; Zhang, X.; Deng, X.; Yu, O.; Zhao, Z.; Gao, X. An adaptive focused target feature fusion network for detection of foreign bodies in coal flow. *Int. J. Mach. Learn. Cybern.* **2023**, *14*, 2777–2791. [CrossRef]
15. Gao, Y.S.; Zhang, B.Q.; Lang, L.Y. Coal and gangue recognition technology and implementation based on deep learning. *Coal Sci. Technol.* **2021**, *49*, 202–208. [CrossRef]
16. Ma, H.W.; Chao, Y.; Xue, X.S.; Mao, Q.H.; Wang, C.W. Binocular vision-based displacement detection method for anchor digging robot. *J. Mine Autom.* **2022**, *48*, 16–25. [CrossRef]
17. Han, Q.; Wang, S.; Fang, Y.; Wang, L.; Du, X.; Li, H.; He, Q.; Feng, Q. A Rail Fastener Tightness Detection Approach Using Multi-source Visual Sensor. *Sensors* **2020**, *20*, 1367. [CrossRef]

18. Wang, H.; Zhang, F.; Wang, H.; Li, Z.; Wang, Y. Real-time detection and location of reserved anchor hole in coal mine roadway support steel belt. *J. Real-Time Image Process.* **2023**, *20*, 89. [CrossRef]
19. Cheng, J.; Wang, D.; Zheng, W.; Wang, H.; Shen, Y.; Wu, M. Position measurement technology of boom-type roadheader based on binocular vision. *Meas. Sci. Technol.* **2024**, *35*. [CrossRef]
20. Wang, Z.; Zhang, G.; Luan, K.; Yi, C.; Li, M. Image-Fused-Guided Underwater Object Detection Model Based on Improved YOLOv7. *Electronics* **2023**, *12*, 4064. [CrossRef]
21. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 1137–1149. [CrossRef]
22. Cheng, L.; Jin, C. X-ray image rotating object detection based on improved YOLOv7. *J. Graphics* **2022**, *44*, 324–334.
23. Chen, R.-C.; Dewi, C.; Zhuang, Y.-C.; Chen, J.-K. Contrast Limited Adaptive Histogram Equalization for Recognizing Road Marking at Night Based on Yolo Models. *IEEE Access* **2023**, *11*, 92926–92942. [CrossRef]
24. Zhang, Z.H.; Wu, Y.; Jin, Z.Q. Detection of Gibberella Infection Rate in Wheat Based on MHSA-YOLOv7. *Radio Eng.* **2024**, *54*, 71–77.
25. Dong, Q.; Wang, Y.L.; Wang, S.K.; Chu, Z.P.; Chen, X.Q.; Yan, S.A. Road Marking Visibility Evaluation Based on Object Detection and Iterative Threshold Segmentation. *J. Tongji Univ. (Nat. Sci.)* **2023**, *51*, 1168–1173+1190.
26. Zhu, L.; Wang, X.; Ke, Z.; Zhang, W.; Lau, R. BiFormer: Vision Transformer with Bi-Level Routing Attention. In Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, BC, Canada, 17–24 June 2023; pp. 10323–10333.
27. Dai, Y.; Liu, W.; Wang, H.; Xie, W.; Long, K. YOLO-former: Marrying YOLO and transformer for foreign object detection. *IEEE Trans. Instrum. Meas.* **2022**, *71*, 1–14. [CrossRef]
28. Zhao, Z.B.; Jiang, Z.G.; Xiong, J.; Nie, L.Q.; Lü, X.C. Fault Classification of Transmission Line Components Based on the Adversarial Continual Learning Model. *J. Electron. Inf. Technol.* **2022**, *44*, 3757–3766.
29. Han, G.; Yang, S.W.; Yuan, P.S.; Zhu, M. Analysis of New Energy Fry Point Defect Detection Based on Improved Faster R-CNN. *Autom. Instrum.* **2023**, *7*, 113–117. [CrossRef]
30. Zhang, Z. A flexible new technique for camera calibration. *IEEE Trans. Pattern Anal. Machine Intell.* **2000**, *22*, 1330–1334. [CrossRef]
31. Wang, H.; Zhang, H.; Deng, H.; Fu, M. Robust monocular visual inertial odometry in radioactive environments using edge-based point features. *Phys. Scr.* **2023**, *98*, 9. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.